



## **Inferring competitive outcomes, ranks and intransitivity from empirical data: A comparison of different methods**

Feng, Yanhao ; Soliveres, Santiago ; Allan, Eric ; Rosenbaum, Benjamin ; Wagg, Cameron ; Tabi, Andrea ; De Luca, Enrica ; Eisenhauer, Nico ; Schmid, Bernhard ; Weigelt, Alexandra ; Weisser, Wolfgang W ; Roscher, Christiane ; Fischer, Markus

**Abstract:** The inference of pairwise competitive outcomes (PCO) and multispecies competitive ranks and intransitivity from empirical data is essential to evaluate how competition shapes plant communities. Three categories of methods, differing in theoretical background and data requirements, have been used: (a) theoretically sound coexistence theory-based methods, (b) index-based methods, and (c) ‘process-from-pattern’ methods. However, how they are related is largely unknown. In this study, we explored the relations between the three categories by explicitly comparing three representatives of them: (a) relative fitness difference (RFD), (b) relative yield (RY), and (c) a reverse-engineering approach (RE). Specifically, we first conducted theoretical analyses with Lotka–Volterra competition models to explore their theoretical linkages. Second, we used data from a long-term field experiment and a short-term greenhouse experiment with eight herbaceous perennials to validate the theoretical findings. The theoretical analyses showed that RY or RE applied with equilibrium data indicated equivalent, or very similar, PCO respectively to RFD, but these relations became weaker or absent with data further from equilibrium. In line with this, both RY and RE converged with RFD in indicating PCO over time in the field experiment as the communities became closer to equilibrium. Moreover, the greenhouse PCO (far from equilibrium) were only similar to the field PCO of earlier rather than later years. Intransitivity was more challenging to infer because it could be reshuffled by even a small competitive shift among similar competitors. For example, the field intransitivity inferred by three methods differed greatly: no intransitivity was detected with RFD; intransitivity detected with RY and RE was poorly correlated, changed substantially over time (even after equilibrium) and failed to explain coexistence. Our findings greatly help the comparison and generalization of studies using different methods. For future studies, if equilibrium data are available, one can infer PCO and multispecies competitive ranks with RY or RE. If not, one should apply RFD with density gradient or time-series data. Equilibria could be evaluated with T tests or standard deviations. To reliably infer intransitivity, one needs high quality data for a given method to first accurately infer PCO, especially among similar competitors.

DOI: <https://doi.org/10.1111/2041-210x.13326>

Posted at the Zurich Open Repository and Archive, University of Zurich

ZORA URL: <https://doi.org/10.5167/uzh-180637>

Journal Article

Accepted Version

Originally published at:

Feng, Yanhao; Soliveres, Santiago; Allan, Eric; Rosenbaum, Benjamin; Wagg, Cameron; Tabi, Andrea; De Luca, Enrica; Eisenhauer, Nico; Schmid, Bernhard; Weigelt, Alexandra; Weisser, Wolfgang W; Roscher, Christiane; Fischer, Markus (2020). Inferring competitive outcomes, ranks and intransitivity from empirical data: A comparison of different methods. *Methods in Ecology and Evolution*, 11(1):117-128.  
DOI: <https://doi.org/10.1111/2041-210x.13326>

**Inferring competitive outcomes, ranks and intransitivity from empirical data: a comparison of different methods**

Yanhao Feng<sup>\*1,2,3</sup>, Santiago Soliveres<sup>4,5</sup>, Eric Allan<sup>5</sup>, Benjamin Rosenbaum<sup>2,6</sup>, Cameron Wagg<sup>7</sup>, Andrea Tabi<sup>7</sup>, Enrica De Luca<sup>7</sup>, Nico Eisenhauer<sup>2,8</sup>, Bernhard Schmid<sup>7</sup>, Alexandra Weigelt<sup>2,9</sup>, Wolfgang W. Weisser<sup>10</sup>, Christiane Roscher<sup>1,2</sup>, Markus Fischer<sup>5</sup>

<sup>1</sup>*UFZ, Helmholtz Centre for Environmental Research, Department of Physiological Diversity, Permoserstraße 15, 04318 Leipzig, Germany*

<sup>2</sup>*German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany*

<sup>3</sup>*State Key Laboratory of Grassland Agro-ecosystems, College of Pastoral Agriculture Science and Technology, Lanzhou University, Lanzhou, 730020, China*

<sup>4</sup>*Department of Ecology, University of Alicante, 03690 Alicante, Spain*

<sup>5</sup>*Institute of Plant Sciences, University of Bern, Altenbergrain 21, 3013 Bern, Switzerland*

<sup>6</sup>*Institute of Biodiversity, University of Jena, Dornburger Strasse 159, 07743 Jena, Germany*

<sup>7</sup>*Department of Evolutionary Biology and Environmental Studies, University of Zurich, Winterthurerstr. 190, CH-8057 Zurich, Switzerland*

<sup>8</sup>*Institute of Biology, University of Leipzig, Deutscher Platz 5e, 04103 Leipzig, Germany*

<sup>9</sup>*Institute of Biology, University of Leipzig, Johannisallee 21, 04103 Leipzig, Germany*

<sup>10</sup>*Terrestrial Ecology Research Group, Department of Ecology and Ecosystem*

*Management, School of Life Sciences Weihenstephan, Technical University of Munich, Hans-Carl-von-Carlowitz-Platz 2, 85354 Freising, Germany*

**\*Author of correspondence:**

Yanhao Feng

Department for Physiological Diversity

26 German Centre for Integrative Biodiversity Research (iDiv) Halle-Jena-Leipzig  
27 Deutscher Platz 5a  
28 04103 Leipzig  
29 Germany  
30 Phone: +49 (0) 341-97-33230  
31 Email: [yanhao.feng@outlook.com](mailto:yanhao.feng@outlook.com)  
32 **Running head:** Inferring competitive outcomes and intransitivity

## 33 **Abstract**

- 34 1. The inference of pairwise competitive outcomes (PCO) and multispecies competitive  
35 ranks and intransitivity from empirical data is essential to evaluate how competition  
36 shapes plant communities. Three categories of methods, differing in theoretical  
37 background and data requirements, have been used: 1) theoretically sound  
38 coexistence theory-based methods, 2) index-based methods, and 3) “process-from-  
39 pattern” methods. However, how they are related is largely unknown.
- 40 2. In this study, we explored the relations between the three categories by explicitly  
41 comparing three representatives of them: 1) relative fitness difference (RFD), 2)  
42 relative yield (RY), and 3) a reverse-engineering approach (RE). Specifically, we first  
43 conducted theoretical analyses with Lotka-Volterra competition models to explore  
44 their theoretical linkages. Second, we used data from a long-term field experiment  
45 and a short-term greenhouse experiment with eight herbaceous perennials to  
46 validate the theoretical findings.
- 47 3. The theoretical analyses showed that RY or RE applied with equilibrium data  
48 indicated equivalent, or very similar, PCO respectively to RFD, but these relations  
49 became weaker or absent with data further from equilibrium. In line with this, both RY  
50 and RE converged with RFD in indicating PCO over time in the field experiment as  
51 the communities became closer to equilibrium. Moreover, the greenhouse PCO (far  
52 from equilibrium) were only similar to the field PCO of earlier rather than later years.  
53 Intransitivity was more challenging to infer because it could be reshuffled by even a  
54 small competitive shift among similar competitors. For example, the field intransitivity  
55 inferred by three methods differed greatly: no intransitivity was detected with RFD;  
56 intransitivity detected with RY and RE was poorly correlated, changed substantially  
57 over time (even after equilibrium) and failed to explain coexistence.

4. Our findings greatly help the comparison and generalisation of studies using different methods. For future studies, if equilibrium data are available, one can infer PCO and multispecies competitive ranks with RY or RE. If not, one should apply RFD with density gradient or time-series data. Equilibria could be evaluated with T tests or standard deviations. To reliably infer intransitivity, one needs high quality data for a given method to first accurately infer PCO especially among similar competitors.

#### **Keywords**

Competitive outcomes, competitive ranks, intransitivity, relative fitness differences, relative yield, a reverse-engineering approach, Chesson's coexistence theory, niche differences

## 68 Introduction

69 The inference of long-term competitive outcomes from empirical data is essential  
70 to understand how competition shapes the structure, dynamics and functioning of plant  
71 communities (Aschehoug *et al.* 2016). For instance, to predict and understand plant  
72 coexistence, we need to explicitly estimate long-term competitive outcomes between  
73 plant species (Hart, Freckleton & Levine 2018), rather than merely the intensity of  
74 competition or overall competitive effects and responses (Aschehoug *et al.* 2016). In  
75 theory, a more competitive species will ultimately exclude a less competitive one over  
76 the long term unless sufficient niche differences exist between the two in space and time  
77 (Gause 1934; Chesson 2000). Therefore, we need to focus on the long-term competitive  
78 outcomes at equilibrium, rather than the short-term outcomes that do not easily  
79 extrapolate to the effects of competition on population dynamics. Based on pairwise  
80 competitive outcomes (PCO), multispecies competitive ranks can be constructed (Keddy  
81 & Shipley 1989). If the ranks are not hierarchical, intransitive competition (i.e.  
82 intransitivity) can emerge as in the “paper-rock-scissors” game: species A beats B which  
83 beats C which in turn beats A (i.e.  $A > B > C > A$ ) (Gallien 2017). The existence of such  
84 intransitivity could promote multispecies coexistence and affect ecosystem functioning  
85 (Levine *et al.* 2017; Maynard, Crowther & Bradford 2017).

86 A number of methods have been developed thus far to infer PCO and thereby  
87 multispecies competitive ranks and intransitivity from empirical data (e.g. de Wit 1960;  
88 Chesson 2000; Ulrich *et al.* 2014; see Table 1 for more examples). Given their  
89 differences in theoretical background and data requirements, we think these methods fall  
90 into three categories (Table 1). The first category of methods are rooted in coexistence  
91 theory, and they infer PCO from parameters in dynamic models that describe long-term  
92 competition between species (MacArthur & Levins 1967; Tilman 1982; Chesson 2000;

see Table 1). These methods provide a theoretically sound measure of PCO, because they indicate which species, in the absence of niche differences, is the competitive winner at equilibrium (Hart, Freckleton & Levine 2018). Among them, relative fitness difference (RFD) derived from Chesson's coexistence theory (Chesson 2000) is the commonly used one in recent studies (e.g. Godoy, Kraft & Levine 2014; Chu & Adler 2015). In theory, RFD is largely consistent with other measures from classical coexistence theories, e.g. MacArthur and Levins (1967)'s  $\frac{K_i}{K_j}$ , May (1974)'s  $\frac{\kappa_i}{\kappa_j}$ , and Tilman (1982)'s  $R^*$  (Carroll, Cardinale & Nisbet 2011; Letten, Ke & Fukami 2017). However, these methods are challenging to apply, because they require a large amount of density gradient or time-series data (often from field experiments) to parameterize dynamic competition models. The second category comprises traditional indices that are often calculated by comparing the performance of a plant of a given species grown in monoculture (or alone) and mixture (Table 1; for a full review, see Weigelt & Jolliffe 2003). Relative yield (RY), i.e. the ratio of a species' yield in mixture to its yield in monoculture (de Wit 1960), is the commonly used among them. The application of RY and other indices (e.g. relative mixture response, relative competition intensity) in this category requires much less data (often from greenhouse experiments). However, these indices have been frequently criticized as they may only indicate short-term rather than long-term competitive outcomes (Freckleton & Watkinson 2000; Hart, Freckleton & Levine 2018). The third category refers to the "process-from-pattern" methods, which often use statistical tools to infer competition from observed species abundance data based on classic community assembly rules (Diamond 1975), e.g. a reverse-engineering approach (RE), C-scores, multivariate logistic regression (Table 1). The RE is a recently developed method that applies Markov chain models to estimate competition matrices that best fit the observed species abundances (Ulrich *et al.* 2014). Within this category,



the RE is the only one that can explicitly estimate PCO and has been frequently applied in recent studies (e.g. Soliveres *et al.* 2015; Maynard *et al.* 2017), but how it relates to methods in the other two categories is not well known (but see Soliveres *et al.* 2018). In short, the three categories of methods not only have different theoretical backgrounds, but also often use data from different sources (e.g. field experiments, greenhouse experiments, observations), which to some extent characterize different conditions of study systems (from equilibrium to non-equilibrium). However, it is largely unknown how different methods are related, which hinders the appropriate interpretation and application of them.

In this study, we chose three commonly used methods: RFD, RY and RE, as representatives of the three categories, respectively, and thoroughly compared them. As it is impossible to comprehensively compare the large number of existing methods (Table 1), and different methods within each category share similarities, we chose specific methods as representatives of different categories to perform a full comparison. First, we used analytical approaches and simulations with Lotka-Volterra competition models to theoretically analyse how the three methods are related to infer PCO and multispecies competitive ranks and intransitivity, when using data from equilibrium to non-equilibrium conditions. Second, we used data from a long-term biodiversity experiment in the field and a short-term microcosm experiment in the greenhouse to validate the theoretical findings. Based on our results, we discuss other methods in the three categories and previous studies using different methods to draw general conclusions, and also provide recommendations for future studies to choose an appropriate method for a particular objective and situation.

## Materials and methods

### THEORETICAL BACKGROUND

#### Three representative methods

##### (1) Relative fitness difference (RFD)

The RFD between two species  $i$  and  $j$  is often defined as  $f_i/f_j = \sqrt{\frac{\alpha_{jj}\alpha_{ji}}{\alpha_{ii}\alpha_{ij}}}$  with Lotka-Volterra competition models (Eqn. 1) (Chesson 2013). Note, however, equivalent measures can also be derived from other phenomenological or mechanistic competition models (e.g. Carroll, Cardinale & Nisbet 2011; Saavedra *et al.* 2017; Hart, Freckleton & Levine 2018).

$$\begin{cases} N_{i,t+1} = N_{i,t} e^{r_i(1 - \alpha_{ii}N_{i,t} - \alpha_{ij}N_{j,t})} & (\text{Eqn. 1a}) \\ N_{j,t+1} = N_{j,t} e^{r_j(1 - \alpha_{jj}N_{j,t} - \alpha_{ji}N_{i,t})} & (\text{Eqn. 1b}) \end{cases}$$

where  $N_{i,t+1}$  and  $N_{i,t}$ , and  $N_{j,t+1}$  and  $N_{j,t}$  are biomasses of species  $i$  and  $j$  at time  $t + 1$  and  $t$ , respectively, and  $r_i$  and  $r_j$  are intrinsic growth rates of species  $i$  and  $j$ , respectively, and alphas ( $\alpha_{ii}$ ,  $\alpha_{jj}$ ,  $\alpha_{ij}$ ,  $\alpha_{ji}$ ) are intraspecific and interspecific competition coefficients.

If  $f_i/f_j > 1$ , species  $i$  will outcompete species  $j$  in the absence of niche differences ( $1 - \sqrt{\frac{\alpha_{ij}\alpha_{ji}}{\alpha_{ii}\alpha_{jj}}}$ ), because species  $j$  is overall more sensitive to competition than species  $i$ . To estimate the parameters in Eqn. 1, data on performance of species at very low densities (to estimate  $r_i$  and  $r_j$ ) and with different densities of conspecific competitors (to estimate  $\alpha_{ii}$  and  $\alpha_{jj}$ ) and heterospecific ones (to estimate  $\alpha_{ij}$  and  $\alpha_{ji}$ )

are required. The required data can be obtained by experimentally creating density gradients (Godoy, Kraft & Levine 2014) or using time-series data of population dynamics with sufficient variation in densities (Chu & Adler 2015).

## Relative yield (RY)

The RY is calculated as a species' yield in mixture divided by its yield in monoculture, often using data from the design of replacement series where total density is held constant (*sensu* de Wit 1960). This does not specify whether the data should be from study systems at equilibrium or not, and in fact RY is often calculated with “non-equilibrium” data from short-term greenhouse experiments (Weigelt & Jolliffe 2003). Species  $i$  is supposed to outcompete species  $j$  if RY of species  $i$  is greater than that of species  $j$  (de Wit 1960), but the theoretical basis for this is rather elusive. Interestingly, RY and related indices are also used to indicate niche differentiation between species, since they appear to reflect the relative strength of intraspecific vs. interspecific competition (by comparing monoculture vs. mixture treatments) (Jolliffe 2000).

## (2) The reverse-engineering approach (RE)

The RE estimates PCO by using Markov chain models to 1) randomly generate a large number (100,000) of competition matrices whose elements  $C_{ij}$  specify the probability that species  $i$  outcompetes species  $j$ , and  $C_{ji} = 1 - C_{ij}$  (if  $C_{ij} > C_{ji}$ , species  $i$  outcompetes species  $j$ ), 2) transform the competition matrices to patch transition matrices from which species abundances can be predicted, and 3) select the competition matrix that best fits the observed abundances of species (for more details, see Ulrich *et al.* 2014). If the goodness-of-fit of the best fitting matrix ( $r_S$ ) is low, it indicates that forces other than competition (e.g. niche-based processes) may also play

important roles in determining species abundances. The RE can be applied to temporal, spatial, or temporal  $\times$  spatial abundance data, and allows one to infer PCO for many species without doing a large number of pairwise competition experiments (e.g. Soliveres *et al.* 2015).

## Multispecies competitive ranks and intransitivity

Using PCOs inferred by any of the methods above, multispecies competitive ranks and intransitivity can be constructed. The competitive ranks can simply be constructed by counting the number of wins each species has. To measure intransitivity, six indices have been proposed to capture different elements of topological variation in an intransitive network (Laird & Schamp 2018) (see Appendix S1 for details): Slater's and Petraitis's  $i$  and Ulrich's  $I$  (overall degree of intransitivity), Kendall and Babington Smith's  $d_s$  (proportion of three-species intransitive loops), Bezembinder's  $\delta$  (proportion of pairs in intransitive relationships), unbeatability ( $u$ ; whether there is an "unbeatable" species, 1, or not, 0), and always-beatability ( $a$ ; whether there is an "always beatable" species, 1, or not, 0).

## THEORETICAL TEST

First, we used the analytical approach with the two-species Lotka-Volterra model (Eqn. 1) to derive the relations between RFD, RY and RE in indicating PCO using data at equilibrium and before equilibrium. To seek generality, we also derived the relations using other competition models (e.g. Beverton-Holt model, annual plant model; see Table S1 in Appendix S2). For the "before-equilibrium" case where it turned out analytical solutions are not possible, however, we used simulations to explore the relations (see below). The details of these theoretical analyses are in Appendix S2.

Second, we used numerical simulations with the Lotka-Volterra model to explore how the relations changed as the systems move away from equilibrium (Appendix S2). In the simulations, we also had eight species and 28 two-species pairs as in the experimental test (see below). Model parameters were randomly drawn from the range of parameters in the Lotka-Volterra models fitted with the field experiment (see below), i.e. with intrinsic growth rates, intra- and inter-specific competition coefficients randomly drawn from 0.2-2, 0.03-0.1 and 0.001-0.05 respectively. We used the drawn competition coefficients to directly calculate RFD. For RY and RE, we first ran the simulations of monocultures and two-species pairs for 100 steps representing 10 years (i.e. 10 steps per year) to be comparable with the field experiment, and recorded biomasses at each step. Equilibrium was achieved around year 4 in most of simulations (see Appendix S2). We then used mean biomass of monocultures and two-species mixtures at three continuous steps around year 1 (before-equilibrium point) and around year 10 (at equilibrium) to calculate RY. For example, for year 1 we used biomasses at steps 0.9, 1 and 1.1 years. Similarly, we applied RE to the time-series (i.e. the three steps) of species relative abundances in two-species mixtures (biomass of each species / total biomass) also around year 1 and 10. Note, although the RE was initially developed for multispecies cases (Ulrich *et al.* 2014), we applied it to two-species cases to make it comparable with the other two methods. The criteria of three methods to indicate if species  $i$  outcompetes species  $j$  are:  $\sqrt{\frac{\alpha_{jj}\alpha_{ji}}{\alpha_{ii}\alpha_{ij}}} > 1$ ,  $RY_i > RY_j$  and  $C_{ij} > C_{ji}$ , respectively. All the simulations were performed with the “deSolve” package (Soetaert, Petzoldt & Setzer 2010) in R 3.4.3 (R Core Team 2018).

Based on PCO inferred by each method, we constructed multispecies competitive ranks for the eight species. For intransitivity, we first theoretically

constructed 30 multispecies communities (species richness>2) as in the experimental test (see below), and then calculated six intransitivity indices for each community based on the PCO. We also counted the number of three-species intransitive loops among the eight species as another measure of intransitivity.

## EXPERIMENTAL TEST

We used data from a long-term field experiment and a short-term greenhouse experiment to validate the findings in the theoretical test. Below we briefly describe the two experiments (for details, see Appendix S3).

### Field experiment

The “dominance” experiment is a long-term grassland biodiversity experiment conducted from 2002 to 2015 in Jena (Roscher *et al.* 2004). The experiment includes nine potentially dominant perennial species: two forbs (*Geranium pratense*, *Anthriscus sylvestris*), two legumes (*Trifolium pratense*, *T. repens*), and five grasses (*Alopecurus pratensis*, *Arrhenatherum elatius*, *Dactylis glomerata*, *Phleum pratense* and *Poa trivialis*). Our analyses excluded *A. sylvestris* due to its absence in the greenhouse experiment. Thus, we used data from 132 plots (1 x 1 m<sup>2</sup>), representing 66 different compositions of communities across five diversity levels (1, 2, 3, 4 and 6 species, with 8, 28, 16, 10 and 4 compositions, respectively), with each composition replicated twice. Species-level annual aboveground biomass was determined each year.

### Greenhouse experiment

A microcosm experiment with the eight species was conducted from April to September in 2011, and repeated in 2012, in a greenhouse in Zurich. An individual of a

species either grew alone (*mono1*), with a conspecific individual (*mono2*), or with a heterospecific individual (*mix*), in pots (11 × 11 × 21 cm<sup>3</sup>). Each treatment was replicated twice, resulting in 88 pots in total. Species-level aboveground biomass was determined in June and September each year.

PCO and multispecies competitive ranks and intransitivity

As in the theoretical test, we applied each method to the field and greenhouse experiments to calculate PCO and multispecies competitive ranks and intransitivity, and count the number of three-species intransitive loops.

(1) RFD

For the field experiment, we first fitted exponential growth models with annual monoculture biomass data from 2002 to 2003, during which all species appeared to experience exponential growth, to estimate their intrinsic growth rates. We then fitted a joint two-species Lotka-Volterra model with time-series biomass data (2002-2015) of all two-species mixtures to simultaneously estimate all intra- and inter-specific competition coefficients. Similarly, for the greenhouse experiment, we estimated intrinsic growth rates with *mono1* biomass data in June and September, and competition coefficients with *mono2* and *mix* data in June and September. After the models for the two experiments were fitted, we used posteriors of competition coefficients to calculate posterior RFD and take their median as an estimate of RFD. To assess the statistical uncertainty of the fits, we calculated a measure of reliability that could range from c. 0.5 to 1: the larger the values are, the more reliable the model fits and estimated RFD are. All the model fittings were performed under a Bayesian framework with the “rstan”

274 package (Stan Development Team 2018) in R 3.4.3 (R Core Team 2018). For more  
275 details of fitting models, see Appendix S4.

## 276 (2) RY

277 For the field experiment, we calculated RY using biomass data from  
278 monocultures and two-species mixtures at each year (2002-2015), but in the main text  
279 we only presented the results of 2015 where communities appeared to be closest to  
280 equilibrium (see Appendix S5 for the results of other years, and equilibrium evaluation  
281 below in Data Analysis). As the monoculture biomass of *P. pratense* in 2015 was zero,  
282 we used its monoculture data from 2014. For the greenhouse experiment, we calculated  
283 RY using *mono1* and *mix* data by averaging the biomasses in June and September.

## 284 (3) RE

285 For the field experiment, we applied RE with data of species relative abundance  
286 (biomass of each species divided by total biomass) of two “spatial” replicates of each  
287 two-species mixture at each year (2002-2015), but similarly here we only presented the  
288 results of 2015 (see Appendix S6 for those of other years). For the greenhouse  
289 experiment, we applied RE with data of two replicates of each two-species mixture.

## 290 DATA ANALYSIS

291 For both theoretical and experimental tests, we tested how PCO (Phi  
292 correlations), multispecies competitive ranks (Wilcoxon signed-rank tests) or values of  
293 each intransitivity index (Phi or Pearson correlations depending whether the variable is  
294 binary or not) inferred by the three methods were related. Moreover, we checked  
295 whether three-species intransitive loops based on different methods differed (Appendix



S7). We also used Pearson correlations to test how each intransitivity index is related to extinction rates (i.e. number of species lost by 2015 divided by total number of species; lower extinction rates indicate greater coexistence) of 30 multispecies communities in the field experiment (Appendix S8). To evaluate whether and when the equilibrium is achieved in the field experiment, we used pairwise T tests to test whether the biomasses of species (in monocultures or two-species mixtures) in the current year were significantly different from those in the previous year (Appendix S9). A significant difference should indicate the system is not at equilibrium. We also used standard deviations to evaluate the equilibrium but the results are similar (for the details, see Appendix S9). We illustrated all the correlations using the 'ggpairs' function in the *GGally* R package (Schloerke *et al.* 2017), and plotted competitive networks of eight species using *igraph* R package (Csardi & Nepusz 2006). All the analyses were performed in R 3.4.3 (R Core Team 2018).

## Results

### PCO

Our analytical and simulation tests demonstrated that, PCO inferred by RY or RE with data from study systems at equilibrium were mathematically equivalent or very similar respectively to those inferred by RFD, regardless of competition models used (Fig. 1a; for details see Appendix S2). This is because  $RY_i - RY_j > 0$  and mostly  $C_{ij} - C_{ji} > 0$  if  $f_i/f_j > 1$  (Appendix S2). However, the simulations showed that these relationships became weaker the further the system was from equilibrium, where  $RY_i - RY_j$  and  $C_{ij} - C_{ji}$  both were affected not only by competition coefficients (as in the equilibrium case) but also by species' intrinsic growth rates and previous biomass or density (Fig. 1a and Appendix S2).

In line with the theoretical findings, our tests with the field experiment showed that PCO inferred by RE (using data from 2015) were strongly positively correlated with those inferred by RFD (Fig. 1b), but for RY its correlation with RFD peaked at 2010 (Fig. S6 in Appendix S5). Moreover, the correlations between PCO inferred by RY or RE and those by RFD became stronger over time (Fig. S6, and Fig. S9 in Appendix S6), as the communities became closer to equilibrium (Appendix S9). Note the equilibrium in monocultures and two-species mixtures was apparently achieved around 2005 and 2008 respectively (Appendix S9). The PCO inferred by the three methods were not correlated in the greenhouse experiment (Fig. 1a), and overall the greenhouse PCO were only positively correlated with those in the field experiment of earlier (2003-2005) rather than later years (Fig. S6 and S9). The reliabilities of the field RFD (mean=0.90 and most >0.80) were much greater than the greenhouse RFD (mean=0.79, with many lower than 0.75) (Figs S4-S5 in Appendix S4).

### 333 Multispecies competitive ranks and intransitivity

334           Multispecies competitive ranks inferred by three methods showed equilibrium vs.  
335 non-equilibrium patterns similar to PCO (Table 2 and Figs 2-3; also see Table S4 in  
336 Appendix S5 and Table S6 in Appendix S6). In the simulations, we found positive  
337 correlations between the three methods with equilibrium data in inferring most of  
338 intransitivity indices (Fig. 4) (also see Appendix S7), but the correlations were overall  
339 weaker than those for PCO with equilibrium data, and also tended to get much weaker  
340 when the system was far from equilibrium (Fig. 4). The tests with the field experiment  
341 showed that there was no intransitivity inferred by RFD (Fig. 5). The RY and RE  
342 detected intransitivity but their intransitivity values (and three-species intransitive loops)  
343 were poorly correlated and also changed substantially over time (Appendix S5-S6).  
344 Moreover, the three methods overall inferred very different values of six intransitivity  
345 indices in the greenhouse experiment (Fig. 5), and the values were also poorly  
346 correlated with the field ones (Fig. 5; Fig. S8 in Appendix S5 and Fig. S11 in Appendix  
347 S6). In addition, all the intransitivity indices were poorly or counterintuitively correlated  
348 with extinction rates (Appendix S8).

## Discussion

Among the three categories, only the coexistence theory-based methods explicitly separate the forces determining competitive outcomes from those causing niche differences, through the strict analysis of population dynamics (Tilman 1982; Chesson 2000). Therefore, they provide a theoretically robust measure of PCO (Hart, Freckleton & Levine 2018), in comparison to the other two categories whose inferences are mainly based on species biomasses or abundances (state variables). Interestingly, however, our theoretical analyses of the three representative methods showed that, RY or RE applied to data at equilibrium, give equivalent or very similar estimates of PCO respectively to RFD. Note, this also implies RY is irrelevant to niche differences in such case, which questions whether RY could, as often claimed (Jolliffe 2000), indicate niche differentiation between species. As different methods within each category share similarities, our findings could to some extent generalize beyond the three chosen representative methods. For example, as RFD can also be derived from a wide range of other competition models (e.g. resource competition model), our findings should therefore also apply when other coexistence theory-based measures (e.g. Tilman's  $R^*$ ) are used to infer PCO. Moreover, most of other index-based methods also compare monoculture vs. mixture treatments as RY (e.g. relative mixture response, relative interaction index), while some compare grown-alone vs. mixture treatments (e.g. relative competition intensity). As the formulas of these indices deviate from that of RY, our theoretical analyses showed that with equilibrium data they could indicate PCO either same to (relative mixture response) or more frequently different from (relative interaction index, relative competition intensity) those inferred by RY and RFD (for details, see Appendix S2). Lastly, among the “process-from-pattern” methods to tease apart the relative importance of competition vs. other forces (such as niche-based processes,

stochasticity) in community assembly, the RE is the only one which can explicitly give competitive outcomes between species. However, our findings of the close link between RE and RFD may also help the other “process-from-pattern” methods (e.g. C-scores, the null model approach) to build their basis of inferring community assembly on species coexistence theory.

In support of the theoretical results, our test with the field experiment showed that both RY and RE converged with RFD in indicating PCO over time as communities got closer to equilibrium. However, for RY the convergence peaked in 2010 rather than 2015, which may be because of the deterioration of monoculture performance for some species after 2010 (Appendix S9), possibly due to the accumulation of pathogens over time (Cortois *et al.* 2016). In line with our findings, a former study found that RY using long-term, presumably equilibrial, data was positively correlated with Tilman’s  $R^*$  (Fargione & Tilman 2006). Moreover, some recent studies also showed that RY at or close to equilibrium could reasonably measure the degree to which Lotka-Volterra models parameterized with monoculture and biculture data capture species abundances of polycultures (Haltz *et al.* 2017; Fort 2018). Our findings also suggest that the former applications of the RE with equilibrium data (e.g. Soliveres *et al.* 2015; Ulrich *et al.* 2016) should reliably indicate PCO. In short, when only the equilibrium data are available, both RY and RE could be a reasonably good proxy of RFD to infer PCO.

In our study, to estimate RFD, we fitted Lotka-Volterra models under a Bayesian framework. For all competition coefficients in a model, we had 10000 posterior values to calculate 10000 posterior RFD ( $\sqrt{\frac{\alpha_{jj}\alpha_{ji}}{\alpha_{ii}\alpha_{ij}}}$ ). If RFD between species  $i$  and  $j > 1$ , species  $i$  is more competitive than species  $j$ , and vice versa. Our “reliability” analyses showed that the estimates of the field RFD were much more reliable than the greenhouse ones,

because for the former on average most of their posterior RFD (90%) agreed with the estimates (either  $> 1$  or  $< 1$ ), a lot greater than that for the latter (79%) (Appendix S4). Moreover, compared to the greenhouse one, the posteriors of competition coefficients in the field experiment yielded much smaller SD relative to means (Appendix S4). These results are likely because the field experiment consists of long time-series data with substantial density variations. However, the data from the greenhouse experiment (and also from most of other greenhouse experiments) have very limited density gradients, and this may have led to less accurate estimation of model parameters.

In contrast to the equilibrium case, both RY and RE only poorly predicted PCOs in non-equilibrium systems. This is because the RY and RE calculated from non-equilibrium data are affected not only by competition coefficients (as in the equilibrium cases) but also by species' growth rates, previous biomass or density (Appendix S2) (Freckleton & Watkinson 2000). This suggests that the use of RY with short-term greenhouse data in many former studies (Table 1) probably fail to correctly indicate PCO. In line with this, our experimental test showed that the greenhouse PCO were only positively correlated with the field PCO from earlier rather than later years. In the non-equilibrium cases, therefore, PCO could only be reliably inferred from RFD by fitting competition models with density gradient or time-series data. However, if the required data are not of high quality (as in our greenhouse experiment), RFD might not make accurate inferences either (Appendix S4), which may explain the poor correlations between the greenhouse and field RFD.

Multispecies competitive ranks inferred by three methods showed similar equilibrium vs. non-equilibrium patterns as PCO. However, competitive relationships among multiple species are more subtle when their competitive ranks are not hierarchical, as intransitivity indices were much more sensitive to the methods used and

the condition of study systems (at equilibrium or not). This is because intransitivity tended to occur mainly among similar competitors ("weak intransitivity") (see Fig. S12 in Appendix S7; Soliveres & Allan 2018). Therefore, even a small shift in competitive outcomes could reshuffle intransitive relationships. In theory, RY or RE with equilibrium data should indicate the same or similar intransitivity as RFD, but our experimental tests showed that the three methods detected very different values of intransitivity. For example, no intransitivity was detected at all with RFD in the field experiment. Both RY and RE detected intransitivity but their values of intransitivity (across all the six indices capturing different topologies of an intransitive network) were poorly correlated and also changed substantially over time (Appendices S5-S6). Moreover, the detected intransitivity failed to explain species coexistence (Appendix S8), which may imply the estimates of intransitivity are not very reliable. These results suggest that the intransitivity inferred by RY or related indices using short-term (often greenhouse) data in former studies (e.g. Keddy & Shipley 1989; Grace, Guntenspergen & Keough 1993; Shipley 1993) are probably not reliable. Moreover, our findings also call for stricter evaluation when inferring intransitivity by RE with equilibrium data (e.g. Soliveres *et al.* 2015) and by RFD with density gradient data (e.g. Godoy *et al.* 2017). In short, all this makes it generally challenging to accurately infer intransitivity regardless of the methods, unless there are high quality data to allow a given method to accurately infer competitive outcomes among similar competitors.

#### Choosing an appropriate method

Based on our findings, we recommend future studies to choose appropriate methods depending on questions of interest and data availability.

- 1) *PCO*. If data of two-species mixtures and monocultures from study systems at or close to equilibrium are available, one can choose RY or RE (or related methods) to infer PCO, because in such cases both RY and RE could reasonably approximate RFD. The application of RE does not need monocultures which may often be unavailable or of low quality and therefore the RE can be more widely applied. Such “equilibrium” data may be available in long-term experiments (e.g. long-term biodiversity experiments) or observations of natural communities. Equilibrium of study systems could be evaluated with methods such as T tests or SD by comparing temporal or spatial samples (Appendix S9). If “equilibrium” data are not available (where RY or RE could not properly indicate PCO), however, one should apply RFD (or related methods) to density gradient or time-series data with sufficient variation in population densities, to accurately estimate PCO.
- 2) *Multispecies competitive ranks and intransitivity*. Our recommendations for inferring multispecies competitive ranks are similar to those for PCO. However, for inferring intransitivity, which is also derived from PCO but very sensitive to even a small shift of PCO among similar competitors, our recommendations are stricter. In other words, to reliably infer intransitivity, one needs high quality data for a given method (e.g. unambiguous equilibrium data for RY and RE, and a large amount of density gradient or time-series data with sufficient density variations for RFD) to first accurately infer PCO especially among similar competitors.



## Acknowledgement

This study has been supported by German Science Foundation (RO2397/8) in the framework of the Jena Experiment (FOR 456/1451). We thank Anne Ebeling for coordination of the field site and contribution of data, as well as the gardeners and numerous students for their help with plot maintenance and during harvest. YHF was also supported by the Fundamental Research Funds for the Central Universities (Izujbky-2019-32). SS was supported by the Spanish Government under a Ramón y Cajal contract (RYC-2016-20604).

## Authors' contributions

YHF and CR, SS, EA, and MF designed the research; CW, AT, NE, BS, AW, WW, CR and MF contributed the data of the field experiment, and EDL and BS contributed the data of the greenhouse experiment; BR parametrized the Lotka-Volterra models with input from YHF. YHF analysed the data; YHF drafted the manuscript, and all other authors contributed substantially to the improvement.

## Data Accessibility

The data of the field and greenhouse experiments associated with this study are archived in the Pangaea (<https://doi.pangaea.de/10.1594/PANGAEA.866358>) and Dryad (<https://doi.org/10.5061/dryad.mgqnk98vz>) respectively.

## 485 References

- 486 Aschehoug, E.T., Brooker, R., Atwater, D.Z., Maron, J.L. & Callaway, R.M. (2016) The  
487 Mechanisms and Consequences of Interspecific Competition Among Plants.  
488 *Annual Review of Ecology, Evolution, and Systematics*, **47**, 263-281.
- 489 Burns, K. (2007) Patterns in the assembly of an island plant community. *Journal of*  
490 *Biogeography*, **34**, 760-768.
- 491 Carroll, I.T., Cardinale, B.J. & Nisbet, R.M. (2011) Niche and fitness differences relate  
492 the maintenance of diversity to ecosystem function. *Ecology*, **92**, 1157-1165.
- 493 Chesson, P. (2000) Mechanisms of Maintenance of Species Diversity. *Annual Review of*  
494 *Ecology and Systematics*, **31**, 343-366.
- 495 Chesson, P. (2013) Species Competition and Predation. *Ecological Systems: Selected*  
496 *Entries from the Encyclopedia of Sustainability Science and Technology* (ed. R.  
497 Leemans), pp. 223-256. Springer New York, New York, NY.
- 498 Chu, C. & Adler, P.B. (2015) Large niche differences emerge at the recruitment stage to  
499 stabilize grassland coexistence. *Ecological Monographs*, **85**, 373-392.
- 500 Cortois, R., Schröder-Georgi, T., Weigelt, A., Putten, W.H. & De Deyn, G.B. (2016) Plant-  
501 soil feedbacks: role of plant functional group and plant traits. *Journal of Ecology*,  
502 **104**, 1608-1617.
- 503 Csardi, G. & Nepusz, T. (2006) The igraph software package for complex network  
504 research. *InterJournal, Complex Systems*, **1695**, 1-9.
- 505 de Wit, C.T. (1960) On competition. Pudoc.
- 506 Diamond, J. (1975) Assembly of species communities. *Ecology and Evolution of*  
507 *Communities* (eds M.L. Cody & J. Diamond), pp. 342-444. Harvard University  
508 Press, Cambridge, Mass, United States.
- 509 Fargione, J. & Tilman, D. (2006) Plant Species Traits and Capacity for Resource  
510 Reduction Predict Yield and Abundance under Competition in Nitrogen-Limited  
511 Grassland. *Functional Ecology*, **20**, 533-540.
- 512 Fort, H. (2018) On predicting species yields in multispecies communities: Quantifying the  
513 accuracy of the linear Lotka-Volterra generalized model. *Ecological modelling*,  
514 **387**, 154-162.
- 515 Freckleton, R. & Watkinson, A. (2000) Designs for greenhouse studies of interactions  
516 between plants: an analytical perspective. *Journal of Ecology*, **88**, 386-391.
- 517 Gallien, L. (2017) Intransitive competition and its effects on community functional  
518 diversity. *Oikos*, **126**, 615-623.
- 519 Gause, G.F. (1934) *The struggle for existence*. The Williams & Wilkins company,  
520 Baltimore.
- 521 Godoy, O., Kraft, N.J. & Levine, J.M. (2014) Phylogenetic relatedness and the  
522 determinants of competitive outcomes. *Ecology Letters*, **17**, 836-844.
- 523 Godoy, O., Stouffer, D.B., Kraft, N.J.B. & Levine, J.M. (2017) Intransitivity is infrequent  
524 and fails to promote annual plant coexistence without pairwise niche differences.  
525 *Ecology*, **98**, 1193-1200.
- 526 Gotelli, N.J. & Graves, G.R. (1996) *Null models in ecology*.
- 527 Grace, J.B., Guntenspergen, G.R. & Keough, J. (1993) The Examination of a  
528 Competition Matrix for Transitivity and Intransitive Loops. *Oikos*, **68**, 91-98.
- 529 Halty, V., Valdés, M., Tejera, M., Picasso, V. & Fort, H. (2017) Modeling plant  
530 interspecific interactions from experiments with perennial crop mixtures to predict  
531 optimal combinations. *Ecological Applications*, **27**, 2277-2289.
- 532 Hart, S.P., Freckleton, R.P. & Levine, J.M. (2018) How to quantify competitive ability.  
533 *Journal of Ecology*, **106**, 1902-1909.

- Jolliffe, P.A. (2000) The Replacement Series. *Journal of Ecology*, **88**, 371-385.
- Keddy, P.A. & Shipley, B. (1989) Competitive Hierarchies in Herbaceous Plant Communities. *Oikos*, **54**, 234-241.
- Laird, R.A. & Schamp, B.S. (2018) Exploring the performance of intransitivity indices in predicting coexistence in multispecies systems. *Journal of Ecology*, **106**, 815-825.
- Letten, A.D., Ke, P.J. & Fukami, T. (2017) Linking modern coexistence theory and contemporary niche theory. *Ecological Monographs*, **87**, 161-177.
- Levine, J.M., Bascompte, J., Adler, P.B. & Allesina, S. (2017) Beyond pairwise mechanisms of species coexistence in complex communities. *Nature*, **546**, 56-64.
- MacArthur, R. & Levins, R. (1967) The Limiting Similarity, Convergence, and Divergence of Coexisting Species. *The American Naturalist*, **101**, 377-385.
- May, R.M. (1974) On the theory of niche overlap. *Theoretical Population Biology*, **5**, 297-332.
- Maynard, D.S., Bradford, M.A., Lindner, D.L., van Diepen, L.T.A., Frey, S.D., Glaeser, J.A. & Crowther, T.W. (2017) Diversity begets diversity in competition for space. *Nature Ecology & Evolution*, **1**, 0156.
- Maynard, D.S., Crowther, T.W. & Bradford, M.A. (2017) Competitive network determines the direction of the diversity-function relationship. *Proceedings of the National Academy of Sciences*, **114**, 11464-11469.
- Miller, T.E., Burns, J.H., Munguia, P., Walters, E.L., Kneitel, J.M., Richards, P.M., Mouquet, N. & Buckley, H.L. (2005) A critical review of twenty years' use of the resource-ratio theory. *The American Naturalist*, **165**, 439-448.
- Ovaskainen, O., Hottola, J. & Siitonen, J. (2010) Modeling species co-occurrence by multivariate logistic regression generates new hypotheses on fungal interactions. *Ecology*, **91**, 2514-2521.
- Perkins, T.A., Holmes, W.R. & Weltzin, J.F. (2007) Multi-species interactions in competitive hierarchies: new methods and empirical test. *Journal of Vegetation Science*, **18**, 685-692.
- R Core Team (2018) R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria.
- Roscher, C., Schumacher, J., Baade, J., Wilcke, W., Gleixner, G., Weisser, W.W., Schmid, B. & Schulze, E.-D. (2004) The role of biodiversity for element cycling and trophic interactions: an experimental approach in a grassland community. *Basic and Applied Ecology*, **5**, 107-121.
- Saavedra, S., Rohr, R.P., Bascompte, J., Godoy, O., Kraft, N.J.B. & Levine, J.M. (2017) A structural approach for understanding multispecies coexistence. *Ecological Monographs*, **87**, 470-486.
- Schloerke, B., Crowley, J., Cook, D., Briatte, F., Marbach, M., Thoen, E., Elberg, A. & Larmarange, J. (2017) GGally: Extension to 'ggplot2'.
- Shipley, B. (1993) A null model for competitive hierarchies in competition matrices. *Ecology*, **74**, 1693-1699.
- Soetaert, K.E., Petzoldt, T. & Setzer, R.W. (2010) Solving differential equations in R: package deSolve. *Journal of Statistical Software*, **33**.
- Soliveres, S. & Allan, E. (2018) Everything you always wanted to know about intransitive competition but were afraid to ask. *Journal of Ecology*, **106**, 807-814.
- Soliveres, S., Lehmann, A., Boch, S., Altermatt, F., Carrara, F., Crowther, T.W., Delgado - Baquerizo, M., Kempel, A., Maynard, D.S. & Rillig, M.C. (2018) Intransitive competition is common across five major taxonomic groups and is

584 driven by productivity, competitive rank and functional traits. *Journal of Ecology*,  
 585 **106**, 852-864.  
 586 Soliveres, S., Maestre, F.T., Ulrich, W., Manning, P., Boch, S., Bowker, M.A., Prati, D.,  
 587 Delgado-Baquerizo, M., Quero, J.L., Schöning, I., Gallardo, A., Weisser, W.,  
 588 Müller, J., Socher, S.A., García-Gómez, M., Ochoa, V., Schulze, E.-D., Fischer,  
 589 M. & Allan, E. (2015) Intransitive competition is widespread in plant communities  
 590 and maintains their species richness. *Ecology Letters*, **18**, 790-798.  
 591 Stan Development Team (2018) RStan: the R interface to Stan. *R package version*  
 592 *2.17.3*. <http://mc-stan.org/>.  
 593 Stone, L. & Roberts, A. (1990) The checkerboard score and species distributions.  
 594 *Oecologia*, **85**, 74-79.  
 595 Tilman, D. (1982) *Resource Competition and Community Structure*. Princeton University  
 596 Press, New Jersey, USA.  
 597 Ulrich, W., Soliveres, S., Kryszewski, W., Maestre, F.T. & Gotelli, N.J. (2014) Matrix  
 598 models for quantifying competitive intransitivity from species abundance data.  
 599 *Oikos*, **123**, 1057-1070.  
 600 Ulrich, W., Zaplata, M.K., Winter, S., Schaaf, W., Fischer, A., Soliveres, S. & Gotelli, N.J.  
 601 (2016) Species interactions and random dispersal rather than habitat filtering  
 602 drive community assembly during early plant succession. *Oikos*, **125**, 698-707.  
 603 Walck, J.L., Baskin, J.M. & Baskin, C.C. (1999) Relative Competitive Abilities and  
 604 Growth Characteristics of a Narrowly Endemic and a Geographically Widespread  
 605 *Solidago* Species (Asteraceae). *American Journal of Botany*, **86**, 820-828.  
 606 Weigelt, A. & Jolliffe, P. (2003) Indices of plant competition. *Journal of Ecology*, **91**, 707-  
 607 720.

608

609 **Table 1** The three categories of methods and their representatives to infer pairwise competitive outcomes and multispecies  
610 competitive ranks and intransitivity.

Category	Theory, representative, related methods	Data requirement	Merits and limitations	Applications
Coexistence theory-based methods	<p><i>Theory:</i> Coexistence theory</p> <p><i>Representative:</i> Relative fitness difference (Chesson 2000)</p> <p><i>Related methods:</i> <math>R^*</math> (Tilman 1982), <math>\frac{K_i}{K_j}</math> (MacArthur &amp; Levins 1967), <math>\frac{\kappa_i}{\kappa_j}</math> (May 1974)</p>	Density gradient or time-series data	<p><i>Merits:</i> theoretically sound</p> <p><i>Limitations:</i> data requirement, model parameterization</p>	<p>Relative fitness difference (e.g. Godoy, Kraft &amp; Levine 2014; Godoy <i>et al.</i> 2017);</p> <p>Related methods (applications of <math>R^*</math> reviewed in Miller <i>et al.</i> 2005)</p>
Index-based methods	<p><i>Theory:</i> theory implicit</p> <p><i>Representative:</i> Relative yield (de Wit 1960)</p> <p><i>Related methods:</i> e.g. relative mixture response, relative interaction index, relative competition intensity, and many other indices (reviewed in Weigelt &amp; Jolliffe 2003)</p>	Monoculture (or grown-alone) and mixture data	<p><i>Merits:</i> easy to apply</p> <p><i>Limitations:</i> theory implicit; not explicitly separating competitive differences from niche differences</p>	<p>Relative yield (e.g. Keddy &amp; Shipley 1989; Walck, Baskin &amp; Baskin 1999; Perkins, Holmes &amp; Weltzin 2007);</p> <p>Related methods (see examples in Weigelt &amp; Jolliffe 2003)</p>
“Process-from-pattern” methods	<p><i>Theory:</i> community assembly rules (Diamond 1975)</p> <p><i>Representative:</i> The reverse-engineering approach (Ulrich <i>et al.</i> 2014)</p> <p><i>Related methods:</i> C-scores (Stone &amp; Roberts 1990), the null model approach (Gotelli &amp; Graves 1996), multivariate</p>	Temporal, spatial or temporal × spatial species abundance data	<p><i>Merits:</i> easy to apply (especially with multispecies communities)</p> <p><i>Limitations:</i> not explicitly separating competitive</p>	<p>The reverse-engineering approach (e.g. Soliveres <i>et al.</i> 2015; Ulrich <i>et al.</i> 2016);</p> <p>Related methods (Gotelli &amp; Graves 1996; Burns 2007; Ovaskainen, Hottola &amp; Siitonen 2010)</p>

	logistic regression (Ovaskainen, Hottola & Siitonen 2010)		differences from niche differences	
--	---	--	------------------------------------	--

611

612 **Table 2** Results of Wilcoxon signed-rank tests of competitive ranks of eight species based on the three different methods (relative  
613 fitness difference: RFD, relative yield: RY, the reverse-engineering approach: RE), in a) the simulation test and b) the experimental  
614 test. In the simulation test, each method is applied to equilibrium (EQM) and before-equilibrium (bef-EQM) conditions. In the  
615 experimental test, each method is applied to the field experiment (field) and the greenhouse experiment (GH). For the field  
616 experiment, both RY and RE used data from 2015. Numbers are the P values of Wilcoxon signed-rank tests.

<b>a)</b>	RFD (EQM)	RY (EQM)	RE (EQM)	RY (bef-EQM)	RE (bef-EQM)
RFD (EQM)		1.000	0.850	0.430	0.730
RY (EQM)			0.850	0.430	0.730
RE (EQM)				0.345	0.526
RY (bef-EQM)					1.000
RE (bef-EQM)					

<b>b)</b>	RFD (field)	RY (field)	RE (field)	RFD (GH)	RY (GH)	RE (GH)
RFD (field)		0.281	0.168	0.393	0.292	0.149
RY (field)			0.546	0.784	1.000	0.386
RE (field)				0.525	0.671	0.230
RFD (GH)					1.000	0.393
RY (GH)						0.202
RE (GH)						

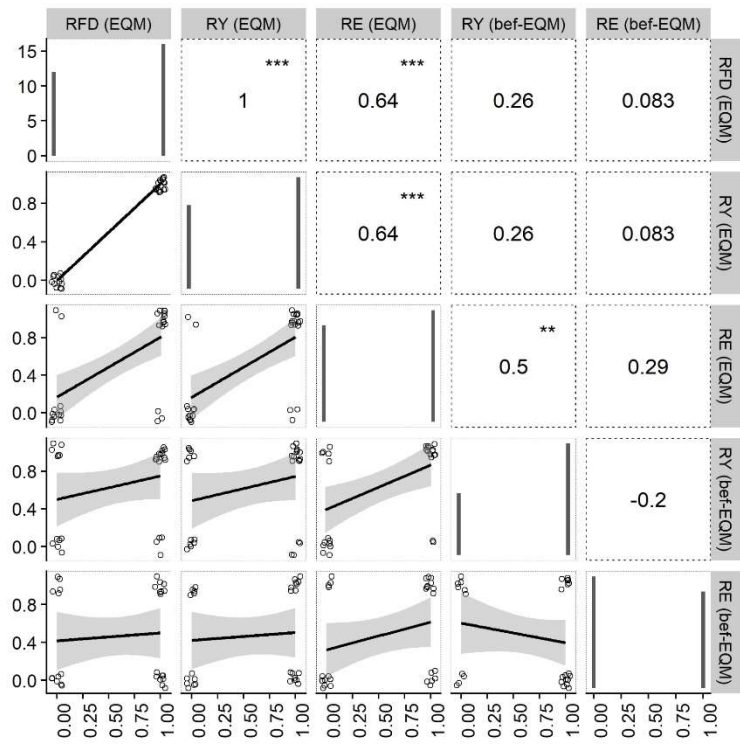
617

618 **Fig. 1** Phi correlations of PCO between eight species (28 pairs in total) based on the three different methods (relative fitness  
619 difference: RFD; relative yield: RY; the reverse-engineering approach: RE), in a) the simulation test and b) the experimental test. In  
620 the simulation test, each method is applied to equilibrium (EQM) and before-equilibrium (before-EQM) conditions. In the experimental  
621 test, each method is applied to the field experiment (field) and the greenhouse experiment (GH). For the field experiment, both RY  
622 and RE used data from 2015. Significance of the correlations is indicated as \*P < 0.05; \*\*P < 0.01; \*\*\*P < 0.001. The lines in the  
623 lower diagonal are based on linear regressions, and the grey area indicates the 95% confidence interval. Data points are slightly  
624 jittered to improve readability.

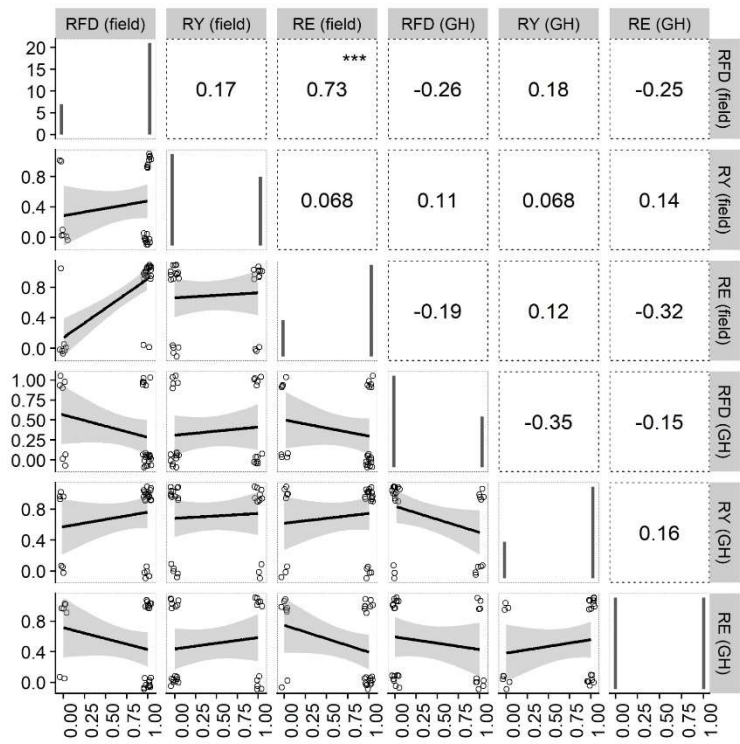
625



a) Simulation test

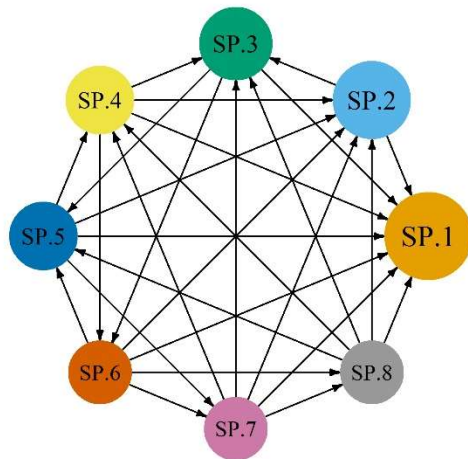


b) Exprimental test

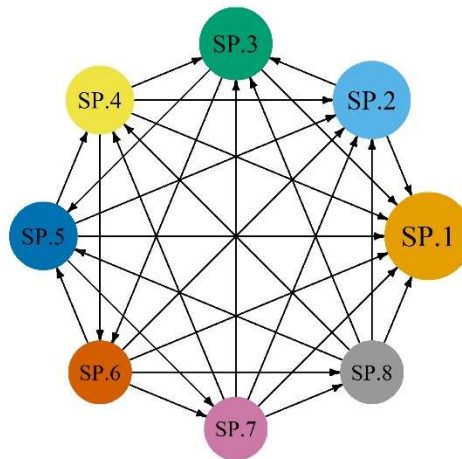


628 **Fig. 2** Competitive ranks and networks of eight simulated species based on the three different methods (relative fitness difference:  
629 RFD; relative yield: RY; the reverse-engineering approach: RE), in the simulation test. Each method is applied to a-c) equilibrium  
630 (EQM) and d-e) before-equilibrium (bef-EQM) conditions. In the networks, node sizes are proportional to species' competitive ranks  
631 (i.e. nodes are larger if ranks are higher), and arrows point to the winner in competition. Different species are indicated with different  
632 colours.

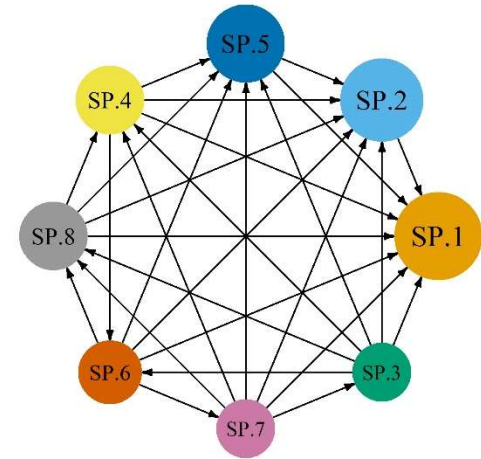
a) RFD (EQM)



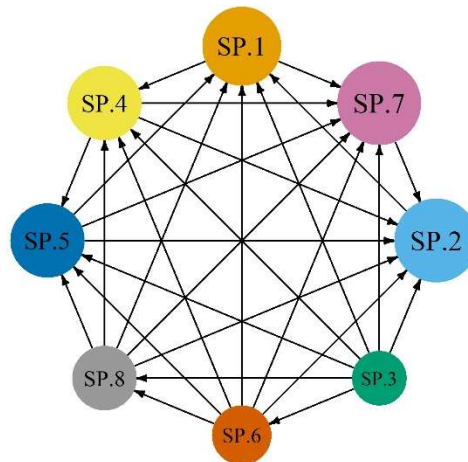
b) RY (EQM)



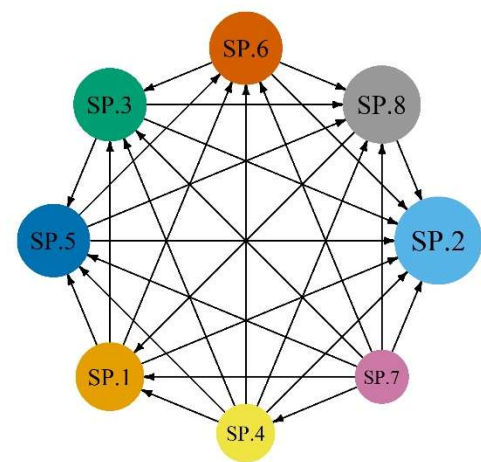
c) RE (EQM)



d) RY (bef-EQM)



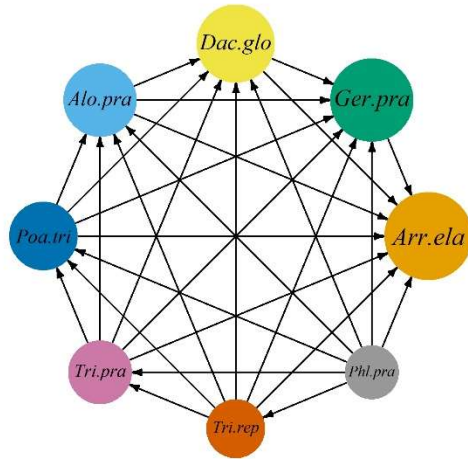
e) RE (bef-EQM)



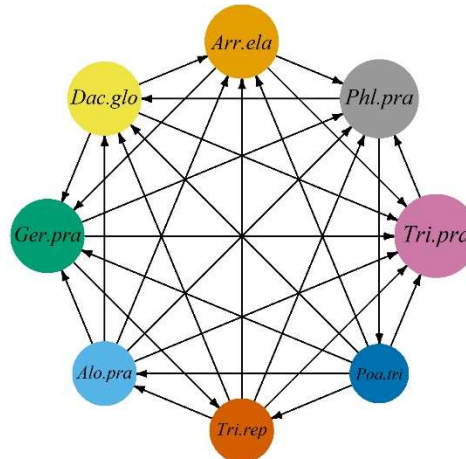
634 **Fig. 3** Competitive ranks and networks of eight species based on the three different methods (relative fitness difference: RFD;  
635 relative yield: RY; the reverse-engineering approach: RE), in the experimental test. Each method is applied to a-c) the field  
636 experiment (field) and d-e) the greenhouse experiment (GH). For the field experiment, both RY and RE used data from 2015. In the  
637 networks, node sizes are proportional to species' competitive ranks (i.e. nodes are larger if ranks are higher), and arrows point to the  
638 winner in competition. Different species are indicated with different colours.

639

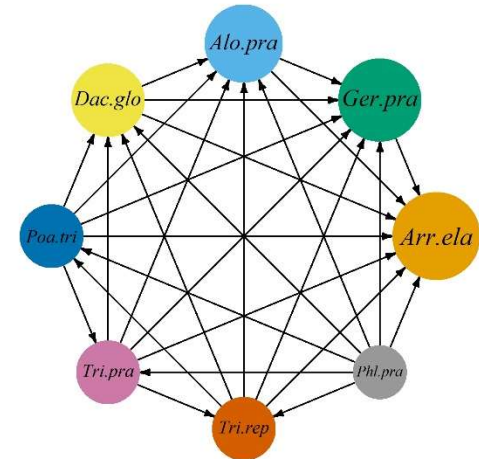
**a) RFD (field)**



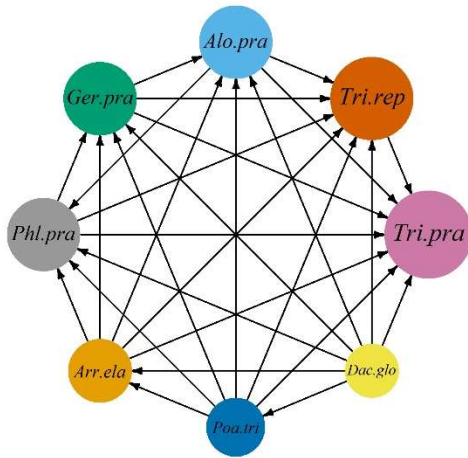
**b) RY (field)**



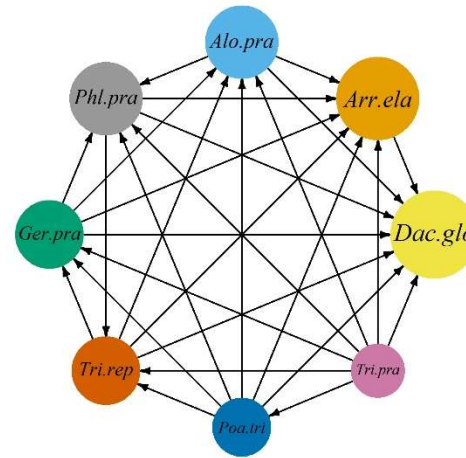
**c) RE (field)**



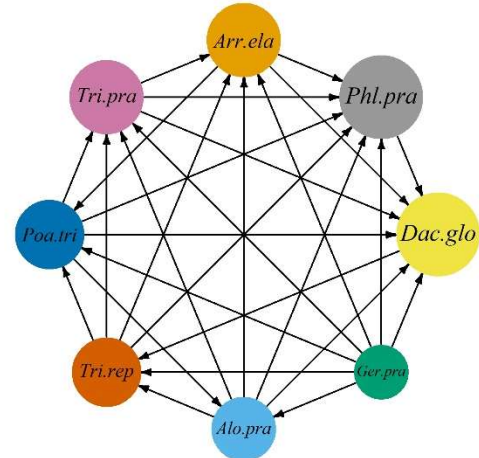
**d) RFD (GH)**



**e) RY (GH)**



**f) RE (GH)**

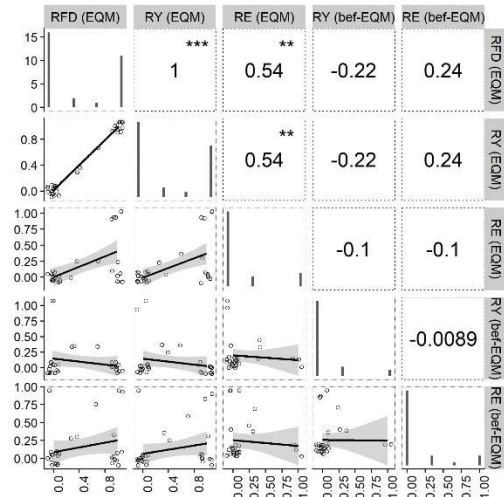


641 **Fig. 4** Pearson correlations of each of intransitivity indices (Slater's and Petraitis's  $i$ , Ulrich's  $I$ , Kendall and Babington Smith's  $d$ ,  
642 Bezembinder's  $\delta$ , unbeatability  $u$  and always-beatability  $a$ ) derived from pairwise competitive outcomes based on each of the three  
643 different methods (relative fitness difference: RFD; relative yield: RY; the reverse-engineering approach: RE), in the simulation test.  
644 Each method is applied to a-c) equilibrium (EQM) and d-e) before-equilibrium (before-EQM) conditions. Significance of correlations is  
645 indicated as \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ . The lines in the lower diagonal are based on linear regressions, and the grey area  
646 indicates the 95% confidence interval. Data points are slightly jittered to improve readability.

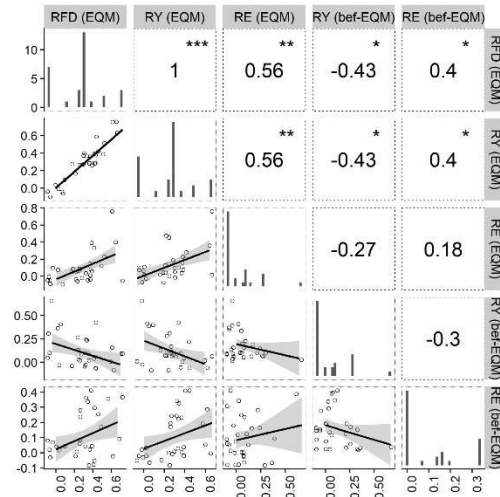
647



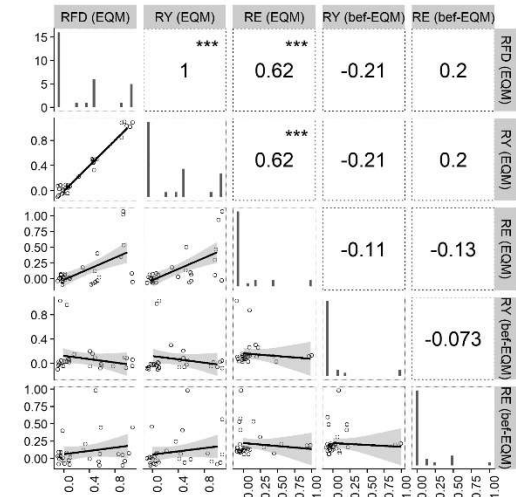
a) Slater's and Petratis's  $i$



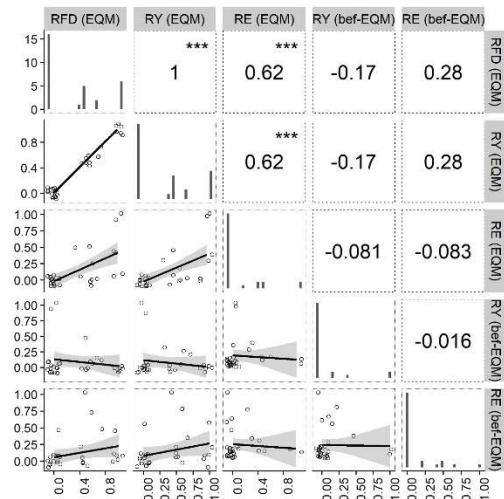
b) Ulrich's  $I$



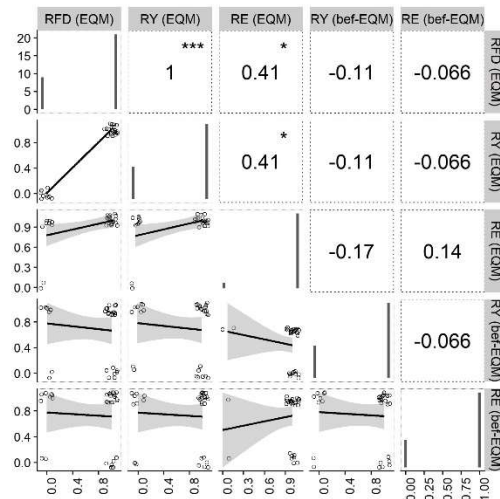
c) Kendall and Babington Smith's  $d_s$



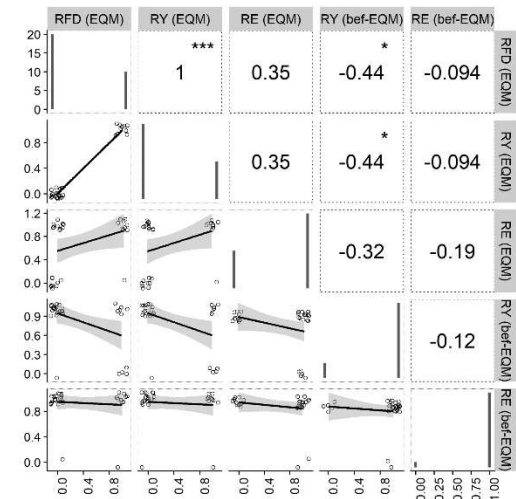
d) Bezembinder's  $\delta$



e) Unbeatability ( $u$ )



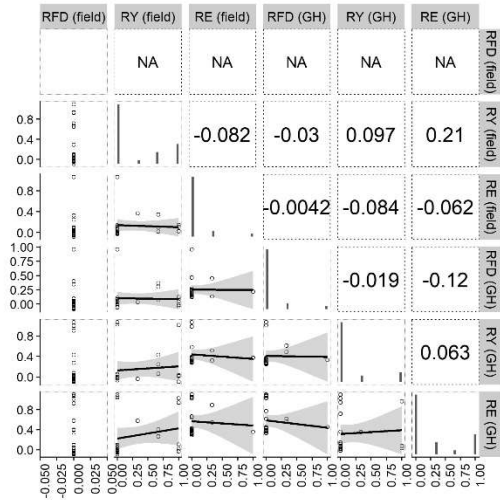
f) Always-beatability ( $a$ )



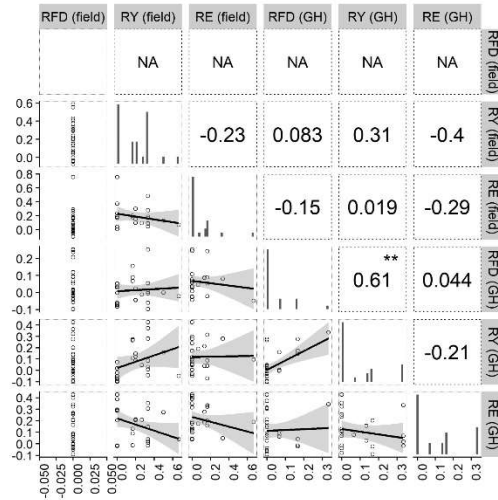
649 **Fig. 5** Pearson correlations of each of intransitivity indices (Slater's and Petraitis's  $i$ , Ulrich's  $I$ , Kendall and Babington Smith's  $d$ ,  
650 Bezembinder's  $\delta$ , unbeatability  $u$ , and always-beatability  $a$ ) derived from the pairwise competitive outcomes based on each of the  
651 three different methods (relative fitness difference: RFD; relative yield: RY; the reverse-engineering approach: RE), in the  
652 experimental test. Each method is applied to the field experiment (field) and the greenhouse experiment (GH). For the field  
653 experiment, both RY and RE used data from 2015, and no intransitivity was detected with RFD so its correlations with others are NA.  
654 Significance of correlations is indicated as \* $P < 0.05$ ; \*\* $P < 0.01$ ; \*\*\* $P < 0.001$ . The lines in the lower diagonal are based on linear  
655 regressions, and the grey area indicates the 95% confidence interval. Data points are slightly jittered to improve readability.



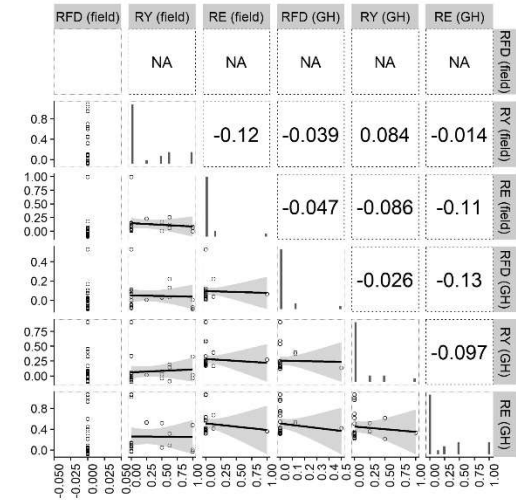
a) Slater's and Petraitis's  $i$



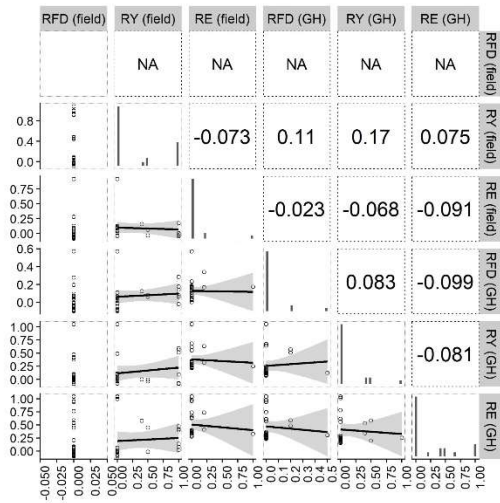
b) Ulrich's  $I$



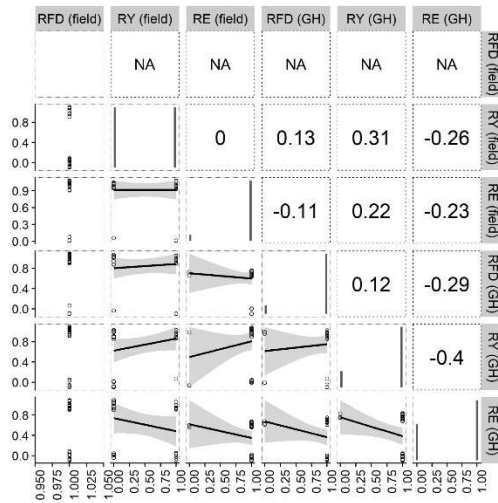
c) Kendall and Babington Smith's  $d_s$



d) Bezembinder's  $\delta$



e) Unbeatability ( $u$ )



f) Always-beatability ( $a$ )

